PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Title: "System and Method for Creating a Data File for Use in Searching a Database"

Applicants: Toshiyuki Nakamura

Attorney Docket No.: JP920020205US1
Serial No.: 10/755,012
Filed: 01/08/2004

Commissioner for Patents
P.O.Box 1450
Alexandria, VA 22313-1450

PRIORITY CLAIM UNDER 35 USC 119(b)

Sir:

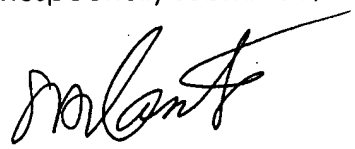
Applicants hereby make a claim for the right of priority of the following non-U.S. patent application:

"Japanese patent application, Serial No. 2003-004572, titled "Database Search System, Search Method Therefor, Method of Creating Data File for Use in Search, and Recording Medium Storing Data File,"

and attach herewith a certified copy thereof.

Date: 05/25/2004

Respectfully submitted,



Samuel A. Kassatly Law Office
20690 View Oaks Way
San Jose, California 95120
Tel. (408) 323-5111
Fax:(408) 323-5112

Samuel A. Kassatly
Attorney for Applicants
Reg. No. 32,247

02 205
SVL

日　本　国　特　許　庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００３年　１月１０日

出願番号
Application Number: 特願２００３－００４５７２

[ST.10/C]: [ＪＰ２００３－００４５７２]

出願人
Applicant(s): インターナショナル・ビジネス・マシーンズ・コーポレーション

２００３年　６月　９日

特許庁長官
Commissioner,
Japan Patent Office 太田信一郎

出証番号　出証特２００３－３０４４８３１

【書類名】　特許願
【整理番号】　JP9020205
【提出日】　平成15年 1月10日
【あて先】　特許庁長官　殿
【国際特許分類】　G06F 17/30
【発明者】
　【住所又は居所】　神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　大和事業所内
　【氏名】　照井　文彦
【発明者】
　【住所又は居所】　神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　大和事業所内
　【氏名】　中村　敏幸
【特許出願人】
　【識別番号】　390009531
　【氏名又は名称】　インターナショナル・ビジネス・マシーンズ・コーポレーション
【代理人】
　【識別番号】　100086243
　【弁理士】
　【氏名又は名称】　坂口　博
【代理人】
　【識別番号】　100091568
　【弁理士】
　【氏名又は名称】　市位　嘉宏
【代理人】
　【識別番号】　100108501
　【弁理士】
　【氏名又は名称】　上野　剛史

【復代理人】

【識別番号】　100104880

【弁理士】

【氏名又は名称】　古部　次郎

【選任した復代理人】

【識別番号】　100118201

【弁理士】

【氏名又は名称】　千田　武

【手数料の表示】

【予納台帳番号】　081504

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9706050

【包括委任状番号】　9704733

【包括委任状番号】　0207860

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　データベース検索システム、その検索方法及び検索に用いられるデータファイルの作成方法並びにデータファイルを格納した記録媒体

【特許請求の範囲】

【請求項１】　文書ファイルを格納した文書データベースと、

前記文書データベースに対する文書ファイルの出し入れを制御するデータベース制御部と、

前記データベース制御部からの検索依頼に応じて、所定の文字列からなるキーワードに基づき前記文書データベースに対する検索を行い、検索結果を前記データベース制御部へ返す検索エンジンと、

前記検索エンジンによる検索処理に用いられ、前記キーワードと当該キーワードの位置情報との対応関係を示す情報を、各キーワードが含まれる文書ファイル内の文書領域に応じて保持するデータファイルと

を備えることを特徴とするデータベース検索システム。

【請求項２】　前記データベース制御部は、前記文書データベースから文書ファイルを読み出して、当該文書ファイルのテキストと当該文書ファイルの構造を示す情報とを抽出して前記検索エンジンに送り、

前記検索エンジンは、前記データベース制御部から受け取ったテキスト及び文書ファイルの構造を示す情報に基づいて、前記データファイルを作成することを特徴とする請求項１に記載のデータベース検索システム。

【請求項３】　前記データファイルは、前記キーワードの位置情報として、キーワードが含まれる文書ファイルを特定する情報及び当該文書ファイルにおける当該キーワードの位置を特定する情報を、前記文書領域ごとに区別される前記キーワードと対応付けて保持することを特徴とする請求項１に記載のデータベース検索システム。

【請求項４】　前記データファイルは、

前記文書データベースに蓄積された文書ファイルに含まれる文字列と当該文字列に関する位置情報へのポインタを、文字列が文書ファイル内の当該文字列が現れる文書領域別に登録した第１のテーブルと、

前記第１のテーブルに登録されている各文字列を含む所定の文字列に関して、当該文字列が存在する文書ファイルを特定する情報及び当該文書ファイルにおける当該文字列の位置を特定する情報を含む位置情報を登録した第２のテーブルとを備えることを特徴とする請求項１に記載のデータベース検索システム。

【請求項５】　コンピュータを用いて文書データベースから所望の文書ファイルを検索するデータベース検索方法であって、

所定の文字列からなる検索タームと当該検索タームが現れる文書ファイル内の文書領域を特定するフィールド情報とを含む検索式を入力する第１のステップと、

所定の文字列からなるキーワードが含まれる文書ファイルを特定する情報と当該キーワードとの対応関係を示す情報を、各キーワードが含まれる文書ファイル内の文書領域に応じて保持する、所定のメモリに格納されたデータファイルを参照し、前記フィールド情報が含まれる前記検索式に対応するキーワードを特定して、特定されたキーワードが含まれる文書ファイルを特定する情報を取得する第２のステップと、

前記データファイルを参照して取得された情報を検索結果として出力する第３のステップと

を含むことを特徴とするデータベース検索方法。

【請求項６】　前記第２のステップでは、前記検索式に含まれる前記フィールド情報に基づき前記キーワードが含まれる文書ファイルの文書領域を区別して、前記文書ファイルを特定する情報を取得することを特徴とする請求項５に記載のデータベース検索方法。

【請求項７】　文書データベースから所望の文書ファイルを検索するために用いられるデータファイルの作成方法であって、

前記文書データベースから文書ファイルを読み出し、各文書ファイルにおけるテキスト及び当該文書ファイルの構造を示す情報とを抽出する第１のステップと、

前記テキストを当該テキストの部分的な文字列からなるキーワードに分割する第２のステップと、

前記文書ファイルの構造を示す情報に基づき、各キーワードが含まれる文書ファイル内の文書領域に応じて、当該キーワードと当該キーワードの位置情報との対応関係を示す情報を保持するデータファイルを作成し、メモリに格納する第３のステップと

を含むことを特徴とするデータファイルの作成方法。

【請求項８】　前記第１のステップでは、前記文書ファイルの構造を示す情報として、当該文書ファイルに記述されたタグの位置を示す情報を抽出することを特徴とする請求項７に記載のデータファイルの作成方法。

【請求項９】　前記第３のステップでは、前記キーワードの位置情報として、キーワードが含まれる文書ファイルを特定する情報及び当該文書ファイルにおける当該キーワードの位置を特定する情報を、前記文書領域ごとに区別される前記キーワードと対応付けて前記データファイルに登録することを特徴とする請求項７に記載のデータファイルの作成方法。

【請求項１０】　文書データベースから所望の文書ファイルを検索するために用いられる索引情報を、コンピュータが読み取り可能に記録した記録媒体であって、

前記文書データベースに蓄積された文書ファイルに含まれる文字列と当該文字列に関する位置情報へのポインタを登録した第１のテーブルと、

前記第１のテーブルに登録されている各文字列を含む所定の文字列に関して、当該文字列が存在する文書ファイルを特定する情報及び当該文書ファイルにおける当該文字列の位置を特定する情報を含む位置情報を登録した第２のテーブルとを備え、

前記第１のテーブルには、同一の文字列が、文書ファイル内の当該文字列が現れる文書領域別に独立に登録され、

前記第２のテーブルには、前記第１のテーブルに登録されている前記文書領域別の各文字列に対応して、当該文書領域に当該文字列が現れる文書を特定する情報及び当該文書ファイルにおける当該文字列の位置を特定する情報が登録されることを特徴とする記録媒体。

【請求項１１】　前記第１のテーブルは、前記文字列を所定の文字コード順

にソートして登録していることを特徴とする請求項１０に記載の記録媒体。

【請求項１２】　前記第１のテーブルには、文書ファイルにおける文書領域に関わらず同一の文字列を１つにまとめた登録がさらに行われ、

前記第２のテーブルには、前記同一の文字列を１つにまとめた登録に対応する当該文字列に関する位置情報の登録がさらに行われていることを特徴とする請求項１０記載の記録媒体。

【請求項１３】　コンピュータを制御して、所定の情報処理を行うプログラムであって、

所定の記憶装置に構築された文書データベースに対する文書ファイルの出し入れを制御するデータベース制御手段と、

所定の文字列からなるキーワードが含まれる文書ファイルを特定する情報と当該キーワードとの対応関係を示す情報を各キーワードが含まれる文書ファイル内の文書領域に応じて保持するデータファイルを参照し、所定の文字列が所定の文書領域に現れる文書ファイルを特定する情報を当該データベース制御手段に送る検索手段として、

前記コンピュータを機能させることを特徴とするプログラム。

【請求項１４】　コンピュータを制御して、所定の情報処理を行うプログラムであって、

所定の文字列からなる検索タームと当該検索タームが現れる文書ファイル内の文書領域を特定するフィールド情報とを含む検索式の入力を受け付ける処理と、

所定の文字列からなるキーワードが含まれる文書ファイルを特定する情報と当該キーワードとの対応関係を示す情報を、各キーワードが含まれる文書ファイル内の文書領域に応じて保持する、所定のメモリに格納されたデータファイルを参照し、前記フィールド情報が含まれる前記検索式に対応するキーワードを特定して、特定されたキーワードが含まれる文書ファイルを特定する情報を取得する処理と、

前記データファイルを参照して取得された情報を検索結果として出力する処理と

を前記コンピュータに実行させることを特徴とするプログラム。

【請求項１５】　コンピュータを制御して、所定の情報処理を行うプログラムであって、

所定のメモリに構築された文書データベースから文書ファイルを読み出し、各文書ファイルにおけるテキスト及び当該文書ファイルの構造を示す情報とを抽出する処理と、

前記テキストを当該テキストの部分的な文字列からなるキーワードに分割する処理と、

前記文書ファイルの構造を示す情報に基づき、各キーワードが含まれる文書ファイル内の文書領域に応じて、当該キーワードと当該キーワードの位置情報との対応関係を示す情報を保持するデータファイルを作成し、所定のメモリに格納する処理と

を前記コンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、データベース検索技術に関し、特に構造化された文書ファイルを格納した文書データベースに対する検索技術に関する。

【０００２】

【従来の技術】

今日、コンピュータを用いたデータベースが広く普及している。その規模も、単体のコンピュータにおいて記憶装置に蓄積されたデータを検索し抽出するものや、インターネット等のネットワーク上に存在する文書ファイルやコンテンツを検索する検索サービスなど、様々である。

【０００３】

インターネットのウェブページに用いられるＨＴＭＬ文書等のように、構造化された文書ファイルでは、文書ファイルにおける部分的な文書領域（以下、フィールド）をタグ等を用いて特定することができ、「タイトル」、「見出し」、「本文」等のフィールドを区別して文書ファイルを作成することが行われている。そこで、この種の文書ファイルを蓄積した文書データベースに対し、所定の文字

列を検索タームとして検索する場合、単にその文字列を含む文書ファイルというだけでなく、文書ファイル中のどのフィールドに当該文字列が含まれるかということまで検索することが行われる（例えば、特許文献１参照）。

【０００４】

【特許文献１】

特開平１０－２９３７６４号公報

【０００５】

【発明が解決しようとする課題】

従来、この種の構造化された文書ファイルに対してフィールドを含む検索を行う検索システムでは、検索タームを含む文書ファイルを検索するための情報（索引）と、各文書ファイルのフィールドの情報とを個別に保持していた。そして、検索の際にこれらの情報を突き合わせて、検索タームが所望のフィールドに含まれている文書ファイルを検索していた。すなわち、まず検索タームを含む文書ファイルを全て検索し、その中から所望のフィールド中に当該文字列を含むものを絞り込む作業が必要であるため、検索に長時間を要していた。

そこで本発明は、フィールド検索を含む文書データベースの検索において、高速な検索処理を実現することを目的とする。

【０００６】

【課題を解決するための手段】

上記の目的を達成する本発明は、次のように構成されるデータベース検索システムとして実現される。すなわち、このデータベース検索システムは、文書データベースと、この文書データベースに対する文書ファイルの出し入れを制御するデータベース制御部と、検索エンジンと、この検索エンジンによる検索処理に用いられ、キーワードとその位置情報との対応関係を示す情報を、各キーワードが含まれる文書ファイル内の文書領域に応じて保持するデータファイルとを備えることを特徴とする。

このデータベースシステムは、単体のコンピュータ装置で実現しても良いし、ネットワークで接続された複数のコンピュータ装置に機能を分けて（例えば別のコンピュータ装置の記憶装置に構築された文書データベースを検索する等）実現

しても良い。

【０００７】

ここで、データファイルは、キーワードの位置情報として、キーワードが含まれる文書ファイルを特定する情報及びこの文書ファイルにおける当該キーワードの位置を特定する情報を、文書領域ごとに区別されるキーワードと対応付けて保持する。より詳しくは、文書データベースに蓄積された文書ファイルに含まれる文字列とこの文字列に関する位置情報へのポインタを、文書ファイル内の文字列が現れる文書領域別に登録した第１のテーブルと、この第１のテーブルに登録されている各文字列を含む所定の文字列（特定の文書領域に属さない文字列を含む）に関して、文字列が存在する文書ファイルを特定する情報及び文書ファイルにおける当該文字列の位置を特定する情報を含む位置情報を登録した第２のテーブルとを備える。

【０００８】

また、上記の目的を達成する他の本発明は、コンピュータを用いて文書データベースから所望の文書ファイルを検索する、次のようなデータベース検索方法としても実現される。すなわち、このデータベース検索方法は、所定の文字列からなる検索タームとこの検索タームが現れる文書ファイル内の文書領域を特定するフィールド情報とを含む検索式を入力する第１のステップと、所定の文字列からなるキーワードが含まれる文書ファイルを特定する情報とキーワードとの対応関係を示す情報を、各キーワードが含まれる文書ファイル内の文書領域に応じて保持する、所定のメモリに格納されたデータファイルを参照し、フィールド情報が含まれる検索式に対応するキーワードを特定して、特定されたキーワードが含まれる文書ファイルを特定する情報を取得する第２のステップと、データファイルを参照して取得された情報を検索結果として出力する第３のステップとを含む。

【０００９】

さらに本発明は、次のような文書データベースから所望の文書ファイルを検索するために用いられる、次のようなデータファイルの作成方法としても実現される。すなわち、文書データベースから文書ファイルを読み出し、各文書ファイルにおけるテキスト及び文書ファイルの構造を示す情報（具体的には、例えば文書

ファイルに記述されたタグの位置を示す情報）とを抽出する第１のステップと、テキストをこのテキストの部分的な文字列からなるキーワードに分割する第２のステップと、文書ファイルの構造を示す情報に基づき、各キーワードが含まれる文書ファイル内の文書領域に応じて、キーワードとキーワードの位置情報との対応関係を示す情報を保持するデータファイルを作成し、メモリに格納する第３のステップとを含む。

【００１０】

また、本発明は、コンピュータを制御して上述したデータベース検索システムにおける各機能を実現し、またはコンピュータに上記のデータベース検索方法あるいはデータファイル（索引ファイル）の作成方法の各ステップに対応する処理を実行させるプログラムとしても実現することができる。このプログラムは、磁気ディスクや光ディスク、半導体メモリ、その他の記録媒体に格納して配布したり、ネットワークを介して配信したりすることにより提供することができる。

【００１１】

【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいて、この発明を詳細に説明する。

データベースには様々な規模や構成のものが存在するが、本実施の形態では、データベース機能を持つアプリケーションプログラムと、当該データベースに対して検索を行う検索エンジンとが動作する、単体のコンピュータ装置で構成されたデータベース検索システムを例として説明する。

【００１２】

図１は、本実施の形態におけるデータベース検索システムを実現するコンピュータ装置のハードウェア構成の例を模式的に示した図である。

図１に示すコンピュータ装置は、演算手段であるＣＰＵ（Central Processing Unit：中央処理装置）１０１と、Ｍ／Ｂ（マザーボード）チップセット１０２及びＣＰＵバスを介してＣＰＵ１０１に接続されたメインメモリ１０３と、同じくＭ／Ｂチップセット１０２及びＡＧＰ（Accelerated Graphics Port）を介してＣＰＵ１０１に接続されたビデオカード１０４と、ＰＣＩ（Peripheral Component Interconnect）バスを介してＭ／Ｂチップセット１０２に接続されたハー

ドディスク１０５、ネットワークインターフェイス１０６及びＵＳＢポート１０７と、さらにこのＰＣＩバスからブリッジ回路１０８及びＩＳＡ（Industry Standard Architecture）バスなどの低速なバスを介してＭ／Ｂチップセット１０２に接続されたフロッピーディスクドライブ１０９及びキーボード／マウス１１０とを備える。

なお、図１は本実施の形態を実現するコンピュータ装置のハードウェア構成を例示するに過ぎず、本実施の形態を適用可能であれば、他の種々の構成を取ることができる。例えば、ビデオカード１０４を設ける代わりに、ビデオメモリのみを搭載し、ＣＰＵ１０１にてイメージデータを処理する構成としても良いし、ＡＴＡ（AT Attachment）などのインターフェイスを介してＣＤ－ＲＯＭ（Compact Disc Read Only Memory）やＤＶＤ－ＲＯＭ（Digital Versatile Disc Read Only Memory）のドライブを設けても良い。

【００１３】

図２は、本実施の形態におけるデータベース検索システムの機能構成を示す図である。

図２を参照すると、本実施の形態のデータベース検索システムは、文書ファイルを蓄積した文書データベース１０と、文書データベース１０に対する文書ファイルの出し入れを制御するデータベース制御部２０と、文書データベース１０に対して検索を行う検索エンジン３０とを備える。本実施の形態のデータベース検索システムを図１に示したコンピュータ装置にて実現した場合、文書データベース１０は、ハードディスク１０５にて実現される。また、データベース制御部２０及び検索エンジン３０は、プログラム制御されたＣＰＵ１０１及びメインメモリ１０３にて実現される。ＣＰＵ１０１を制御するプログラムは、磁気ディスクや光ディスク、半導体メモリ、その他の記録媒体に格納して配布したり、ネットワークを介して配信したりすることにより提供される。図２に示したコンピュータ装置では、このプログラムがハードディスク１０５に保存（インストール）された後、メインメモリ１０３に読み込まれ展開されて、ＣＰＵ１０１を制御し、上記の各機能を実現させる。

なお、特に図示しないが、データベース検索システムは、文書データベース１

０に格納する文書ファイルや検索条件を指定する検索式、検索処理を要求するコマンド等の入力を行うための入力手段（例えば図１に示したキーボード／マウス１１０）を備えることができる。また、検索結果を出力する表示手段（ディスプレイ装置等）を備えることができる。本実施の形態のデータベース検索システムを、インターネット等のネットワーク上の検索サーバとして構築した場合は、これら入出力手段として、例えば図１に示したネットワークインターフェイス１０６を用い、ネットワークを介して接続された端末装置からの入力を受け付け、当該端末装置に検索結果を返すといった実施態様も可能である。

【００１４】

上記の構成において、文書データベース１０は、ＨＴＭＬ文書等の構造化された文書ファイルを格納している。格納された文書ファイルは、フィールドに分けて文書を記述している。なお、フィールドの構成ルールは文書ファイルを記述した言語に応じて任意であり、所定のフィールド内にさらに下位のフィールドを設けて多重化することもできるし、文書ファイル全体を１つのフィールドとすることもできる。また、文書構造（フィールドの構成）の記述方法も、ＨＴＭＬ文書等で用いられるように文書中にタグを埋め込むことで記述する他、テキストデータとフィールド位置を示すデータとをセットにしてファイル化する等、種々の方法を用いて記述することができる。以下では、タグを用いて文書構造を記述する形式を例として説明する。

【００１５】

データベース制御部２０は、文書データベース１０に対して文書ファイルの格納及び読み出しを行う。所望の文書ファイルを読み出す際には、検索ターム及びフィールドを特定する情報（以下、フィールド情報）を含む検索式を検索エンジン３０に渡し、得られた検索結果を用いて所望の文書ファイルを特定し、文書データベース１０から読み出す。また、文書ファイルの読み出し処理に影響しない所定のタイミングで、検索エンジン３０が文書ファイルの検索に使用する情報テーブルである索引ファイル３１を作成するための情報を、文書データベース１０から読み出して検索エンジン３０に渡す。検索時及び索引ファイル３１作成時の動作や検索式の詳細については後述する。

【００１６】

検索エンジン３０は、データベース制御部２０からの要求に応じて、検索式に基づき文書データベース１０の文書ファイルを検索する。検索は、索引ファイル３１を参照して行われる。索引ファイル３１は、文書ファイルに含まれる文字列（キーワード）と当該文書ファイルにおける当該文字列の位置の情報とを対応付けたデータファイルであって、検索エンジン３０による検索処理に先立って予め作成され、例えば図１のメインメモリ１０３に格納されており、検索エンジン３０による検索処理において使用される。

検索式に該当する文書ファイルが見つかった（ヒットした）ならば、当該文書ファイルに関する情報として、文書ファイルを特定する文書番号と、当該文書ファイルにおける検索タームに対応する文字列の位置の情報とをデータベース制御部２０に返す（当然ながら、検索式に該当する文書ファイルが見つからなかった（ヒットしなかった）場合、ヒットしなかったことを通知するメッセージが返される）。これにより、データベース制御部２０は、所望の文書ファイルを指定して文書データベース１０から読み出すことができる。検索エンジン３０による検索処理の詳細については後述する。

【００１７】

索引ファイル３１は、検索のためのキーワードと当該キーワードの存在位置を示す位置情報へのポインタとを登録したテーブルであるキーファイル３２と、キーファイル３２に登録されているキーワードが存在する文書ファイルを特定する情報及び当該文書ファイルにおける当該キーワードの位置情報を登録したテーブルであるＰＯＳファイル（位置情報ファイル）３３とで構成される。索引ファイル３１の作成は、データベース制御部２０による文書ファイルの読み出し処理及びその中で検索エンジン３０に依頼される検索処理に影響しない所定のタイミングで行われる。索引ファイル３１の構成については、後に図７を用いてさらに詳細に説明する。

【００１８】

ここで、本実施の形態による検索の大まかな処理の流れを説明する。

図３は、データベース制御部２０と検索エンジン３０との間のデータのやり取

りを示す図であり、図３（Ａ）は索引ファイル３１の作成時の様子、図３（Ｂ）は検索時の様子をそれぞれ示す。

文書の検索を行うには、検索対象となる文書ファイルの情報に基づいて、予め索引ファイル３１を作成しておく必要がある。ＨＴＭＬ文書のように構造を持った文書ファイルを索引ファイル３１に登録する場合、まず、データベース制御部２０により、登録対象である文書ファイルから、タグを取り除いた文書データ（以下、テキスト）と、各タグが付加されていた文書ファイル中の位置の情報とが抽出される（図３（Ａ）参照）。この際、後の検索時に所望の文書ファイルを抽出しやすいように、元の文書ファイルに独自の情報を追加することも可能である。各文書ファイルには文書番号が重複しないように割り当てられ、テキスト、タグの位置情報及び文書番号の各情報がデータベース制御部２０から検索エンジン３０に送られる。また、各文書ファイルにおけるフィールドの位置情報もデータベース制御部２０から検索エンジン３０に送られる。

【００１９】

検索エンジン３０は、テキスト内の文字列を、単語（可変長連鎖）もしくは所定数の文字の連鎖（固定長連鎖）に切り分け、これらの連鎖がテキストの何文字目に現れるかを示す情報（文字位置規則）を、当該テキストの文書番号と共に索引ファイル３１に登録しておく。この索引の手法は、一般に転置索引として周知である。また、文書ファイルにおけるフィールドの位置情報を、転置索引において切り分けられた連鎖と同様の文字位置規則に変換して登録しておく。本実施の形態では、さらに、このフィールドの位置情報を上述の連鎖の各々に付随させておく。

【００２０】

検索処理を実行する場合、まずデータベース制御部２０において検索式が生成される。検索式は、検索タームと共に、必要に応じてフィールド情報を含む。フィールド情報を検索式に含むことにより、所望のフィールドに検索タームを含む文書ファイル（例えば、タイトルに「日本」という文字列を含む文書ファイル）を指定して検索することが可能である。もちろん、文書ファイルのどこかに検索タームが含まれるものを検索するのであれば、検索式においてフィールド情報を

指定しないこともできる。生成された検索式は、データベース制御部２０から検索エンジン３０に送られる（図３（Ｂ）参照）。

【００２１】

検索式が与えられると、検索エンジン３０は、当該検索式を解釈し、索引ファイル３１を参照して、当該検索式を満たす文書ファイルの文書番号と、当該検索式の検索タームである文字列がテキストの何文字目に現れるかを示す文字位置情報とを取得し、データベース制御部２０に返送する。なお、検索式を満たす文書ファイルが存在しなかった場合は、これを通知するメッセージが返送されることとなる。

本実施の形態では、上述したようにフィールドの位置情報を文字列（連鎖）の各々に付随させたことにより、特定のフィールドに含まれる特定の文字列を検索しようとする場合に、高速な処理を実現する。

データベース制御部２０は、検索エンジン３０から文書番号と文字位置情報とを受け取り、これらの情報に基づいて、文書データベース１０から所望の文書ファイルを読み出すことができる。

【００２２】

次に、検索エンジン３０による検索手法について詳細に説明する。

検索エンジン３０は、上述したように、検索ターム及びフィールド情報を含む検索式をデータベース制御部２０から受け取り、当該検索タームを含む文書ファイルを特定する文書番号と、当該文書ファイルにおける検索タームに対応する文字列の位置の情報とをデータベース制御部２０に返す。この文字列の位置の情報には、文書ファイルにおけるフィールドの情報も含まれる。すなわち、検索エンジン３０は検索式に応じたフィールドの検索も実行する。

以下では、説明の便宜上、まずフィールド検索に触れず、文書ファイルのテキストから所望の文字列を検索する方法について説明し、次いでフィールド検索の方法を説明することとする。

【００２３】

検索エンジン３０による文字列の検索手法としては、従来から知られている任意の手法を用いることができるが、本実施の形態では、ｎ－ｇｒａｍモデルを用

いた解析による手法を例として説明する。

まず、日本語文等のように、文を記述する際に単語による区切りの表れないテキストから所望の文字列を検索する場合について説明する。この場合、テキストは固定長ずつの連鎖に区切られ、キーワードとして、索引ファイル３１に登録される。以下、具体例を挙げて説明する。

【００２４】

「明日は明日の風が吹く。」

というテキストから所望の文字列を検索する場合を考える。検索エンジン３０において、当該テキストの索引ファイル３１への登録が、次のように行われる。

まず、テキストをｎ文字ずつの文字連鎖（以下、キーワードと呼ぶ）に分解する。例としてｎ＝２とすると、「明日は明日の風が吹く。」は、次のように分解される。

明日
日は
は明
明日
日の
の風
風が
が吹
吹く
く。

【００２５】

各キーワードに関し、先頭のキーワードの位置番号を「１」として１文字ずれるごとに１つずつ位置番号を増やしていく。この文書ファイルの文書番号を「０」として（文書番号、位置番号）のように表すと、次のようになる。

明日　　（０，１）
日は　　（０，２）
は明　　（０，３）
明日　　（０，４）
日の　　（０，５）
の風　　（０，６）
風が　　（０，７）
が吹　　（０，８）
吹く　　（０，９）
く。　　（０，１０）

これを、各キーワードの文字コード（ＡＳＣＩＩ、ＪＩＳコード、Ｕｎｉｃｏｄｅ等）順でソートすると、次のようになる。

く。　　（０，１０）
の風　　（０，６）
は明　　（０，３）
が吹　　（０，８）
日の　　（０，５）
日は　　（０，２）
明日　　（０，１）
明日　　（０，４）
風が　　（０，７）
吹く　　（０，９）

【００２６】

以上の情報のうち、キーワードがキーファイル３２に登録され、文書番号及び位置番号がＰＯＳファイル３３に登録される。同一のキーワードは、キーファイル３２には１つしか登録されないが、対応する複数の文書番号及び位置番号の組

（位置情報）がＰＯＳファイル３３に登録される。

図４は、上記のテキストに対する索引ファイル３１の構成を示す図である。

【００２７】

次に、検索時の動作について説明する。

検索ターム「明日の風」を含む検索式が、データベース制御部２０から検索エンジン３０に送られたものとする。この場合、検索エンジン３０は、まず検索ターム「明日の風」を２文字ずつ区切り、「明日」と「の風」とする。そして、これらに対応するキーワードの位置情報を、索引ファイル３１から取得する。図４に示した索引ファイル３１によれば、キーワード「明日」の位置情報は、（０，１）及び（０，４）であり、キーワード「の風」の位置情報は、（０，６）である。これらの位置情報を参酌すると、（０，４）の位置の「明日」と（０，６）の位置の「の風」が連続していることが分かり、結果として（０，４）の位置に存在する「明日の風」という文字列を検索結果としてデータベース制御部２０へ返すことができる。

【００２８】

次に、英語文等のように、文を記述する際に単語による区切りが現れるテキストから所望の文字列を検索する場合について説明する。この場合、各単語の文字列としての長さは様々であるため、かかる可変長連鎖である単語をそのまま索引ファイル３１のキーワードとすると、キーファイル３２内から求めるキーワードを探すのが困難となる。そこで、この可変長連鎖を固定長連鎖に分解して検索を行うための機構が導入される。以下、具体例を挙げて説明する。

【００２９】

「to be or not to be that is the question」

というテキストから所望の文字列を検索する場合を考える。当該テキストの索引ファイル３１への登録が、次のように行われる。

単純に文字列中の空白（スペース）を単語の区切りとして分解し、この文書の文書番号を「１」とすると、次のようになる。

ｔｏ　　（１，１）

ｂｅ	（１，３）
ｏｒ	（１，５）
ｎｏｔ	（１，７）
ｔｏ	（１，１０）
ｂｅ	（１，１２）
ｔｈａｔ	（１，１４）
ｉｓ	（１，１８）
ｔｈｅ	（１，２０）
ｑｕｅｓｔｉｏｎ	（１，２３）

これを、各キーワードの文字コード順でソートし、重複するキーを１つにまとめて、索引ファイル３１に登録される。

図５は、上記のテキストに対する索引ファイル３１の構成を示す図である。

【００３０】

可変長連鎖に対する索引ファイル３１では、検索効率を高めるため、さらに次のような関係ファイル３４が作成される。

まず、各単語に単語の開始マーク（表記上は＾で表す）と終了マーク（表記上は＄で表す）とを付し、マークの付された単語をｎ文字ごとに区切る。例としてｎ＝３とし、単語「ｑｕｅｓｔｉｏｎ」を区切ると、文字列「＾ｑｕｅｓｔｉｏｎ＄」は、次のように分解される。

＾ｑｕ
ｑｕｅ
ｕｅｓ
ｅｓｔ
ｓｔｉ
ｔｉｏ
ｉｏｎ
ｏｎ＄

すなわち、ｎ文字の固定長のキーワードの集まりとして単語を表現したことになる。ここで、キーワードに対する位置情報（ＰＯＳ）にあたる情報を、（キーワード番号、単語内位置番号）として定義する。キーワード「ｑｕｅｓｔｉｏｎ」のキーワード番号を「４」とすると、上記の各文字列に対して次のような情報が付加される。

```
^ｑｕ　（４，１）
　ｑｕｅ　（４，２）
　　ｕｅｓ　（４，３）
　　　ｅｓｔ　（４，４）
　　　　ｓｔｉ　（４，５）
　　　　　ｔｉｏ　（４，６）
　　　　　　ｉｏｎ　（４，７）
　　　　　　　ｏｎ＄　（４，８）
```

これらの文字列及び位置情報を、キーファイル３２及びＰＯＳファイル３３と同様に、文字コード順でソートして関係ファイル３４に登録する。

図６は、「＾ｑｕｅｓｔｉｏｎ＄」に対する関係ファイル３４の構成及びキーファイル３２との関係を示す図である。

【００３１】

次に、検索時の動作について説明する。

検索ターム「ｑｕｅｓｔｉｏｎ」を含む検索式が、データベース制御部２０から検索エンジン３０に送られたものとする。この場合、検索エンジン３０は、まず検索タームの文字列に開始マーク及び終了マークを付した「＾ｑｕｅｓｔｉｏｎ＄」をｎ文字ごとの連鎖に分解し、次の文字列のセットを得る。

```
^ｑｕ
ｅｓｔ
```

ｉｏｎ

ｏｎ＄

そして、関係ファイル３４を参照し、文字列「＾ｑｕ」が１文字目、文字列「ｅｓｔ」が４文字目、文字列「ｉｏｎ」が７文字目、文字列「ｏｎ＄」が８文字目に現れるキーワードを探す。すると、図６に示したキーワード「ｑｕｅｓｔｉｏｎ」の関係ファイル３４において、文字列「＾ｑｕ」の位置情報が（４，１）、文字列「ｅｓｔ」の位置情報が（４，４）、文字列「ｉｏｎ」の位置情報が（４，７）、文字列「ｏｎ＄」の位置情報が（４，８）である。したがって、キーワード番号「４」のキーワードが「ｑｕｅｓｔｉｏｎ」であることが分かる。

この検索結果に基づき、キーファイル３２及びＰＯＳファイル３３を参照し、単語「ｑｕｅｓｔｉｏｎ」の位置情報が（１，２３）であることが分かる。したがって、文書番号１番の文書ファイルにおけるテキストの２３文字目に、検索タームにかかる単語が現れることが分かる。

【００３２】

次に、フィールド検索の方法について説明する。

本実施の形態は、このフィールド検索において顕著な特徴を有するので、まず本実施の形態による検索方法の概念を説明した後、従来のフィールド検索の手法と対比して具体的な動作例を説明することとする。

図７は、本実施の形態による索引ファイル３１を用いたフィールド検索の方法を説明する図である。

本実施の形態では、キーファイル３２において、各キーワードに関して、当該キーワードが現れるフィールドを区別して登録する。図７に示す例では、キーワード「ｑｕｅｓｔｉｏｎ」は、それ自身が登録されると共に、フィールドＦ１に現れる「ｑｕｅｓｔｉｏｎ」、フィールドＦ２に現れる「ｑｕｅｓｔｉｏｎ」、フィールドＦ３に現れる「ｑｕｅｓｔｉｏｎ」等が個別のアイテムとして登録されている。

一方、ＰＯＳファイル３３においては、キーワードの位置情報と共に、各フィールドに現れる当該キーワードの位置情報を登録する。図７に示す例では、キー

ワード「ｑｕｅｓｔｉｏｎ」の位置情報と、フィールドＦ１に現れる「ｑｕｅｓｔｉｏｎ」の位置情報、フィールドＦ２に現れる「ｑｕｅｓｔｉｏｎ」の位置情報等がそれぞれ登録されている。

【００３３】

索引ファイル３１を以上のように構成してフィールドの位置情報をキーワードに付随させておくことにより、特定のフィールドに現れる特定の文字列を指定して検索しようとする場合、キーファイル３２の該当アイテムからＰＯＳファイル３３を参照することにより、直ちに所望の文字列の位置情報を得ることができる。図７に示す例では、例えばフィールドＦ１に現れる「ｑｕｅｓｔｉｏｎ」を検索しようとする場合、検索タームとして「ｑｕｅｓｔｉｏｎ」、フィールドとしてＦ１を指定すれば、キーファイル３２の該当アイテムからＰＯＳファイル３３を直接参照し、（Ｄｏｃ１５，Ｐｏｓ１１）、（Ｄｏｃ３２，Ｐｏｓ１３）、（Ｄｏｃ９５，Ｐｏｓ２５）といった文書番号及び位置番号の組が直ちに得られる。

【００３４】

次に、フィールド検索の動作を、具体例を挙げて説明する。

次に示す文書ファイルから所望の文字列を検索する場合を考える。

<title>IBM software</title>
This page explains IBM software products

上記の文書ファイルは、タグによって構造化されており、<title>タグで囲まれた範囲をフィールド１と定義することとする（なお、ＩＢＭは米国ＩＢＭ社の商標）。

【００３５】

まず、データベース制御部２０から検索エンジン３０へ、文書ファイルからタグが取り除かれたテキスト「IBM software This page explains IBM software products」と、<title>タグの位置情報とが送られる。そして、検索エンジン３０において、この文書ファイルに関する情報が索引ファイル３１に登録される。

このテキストに関して、文書番号を「2」とし、上記と同様の方法でキーワードとその位置情報を抽出すると、登録情報は次のようになる。

ＩＢＭ	（２，１）
ｓｏｆｔｗａｒｅ	（２，４）
Ｔｈｉｓ	（２，１２）
ｐａｇｅ	（２，１６）
ｅｘｐｌａｉｎｓ	（２，２０）
ＩＢＭ	（２，２８）
ｓｏｆｔｗａｒｅ	（２，３１）
ｐｒｏｄｕｃｔｓ	（２，３９）

【００３６】

また、フィールド１の定義は、タグ位置に基づき、開始位置がキーワード「ＩＢＭ」で、終了位置がキーワード「ｓｏｆｔｗａｒｅ」となっている。そこで、この位置情報を索引ファイル３１に登録する。位置番号は、開始位置が、「１」であり、終了位置が、キーワード「ｓｏｆｔｗａｒｅ」の次の位置になるので「１２」である。したがって、フィールド１に関する登録情報は次のようになる。

（フィールド１）　　（２，１）、（２，１２）

【００３７】

従来の検索エンジン３０では、単にこのフィールド１に関する位置情報を各キーワードの位置情報と共に索引ファイル３１に登録していた。これに対し、本実施の形態では、フィールド１に関する情報を各キーワードに付随させて登録する。具体的には、キーワードとそのキーワードが現れるフィールドとを組み合わせ、フィールドに組み合わされたキーワードを独立のキーワードとして登録する。この操作により、登録情報は次のようになる。

ＩＢＭ	（２，１）
ＩＢＭ（Ｆ１）	（２，１）
ｓｏｆｔｗａｒｅ	（２，４）
ｓｏｆｔｗａｒｅ（Ｆ１）	（２，４）
Ｔｈｉｓ	（２，１２）
ｐａｇｅ	（２，１６）
ｅｘｐｌａｉｎｓ	（２，２０）
ＩＢＭ	（２，２８）
ｓｏｆｔｗａｒｅ	（２，３１）
ｐｒｏｄｕｃｔｓ	（２，３９）

これを文字コード順でソートし、重複するキーワードをまとめて索引ファイル３１に登録される。

図８は、上記の文書ファイルに対して最終的に得られる、本実施の形態における索引ファイル３１の構成を示す図である。

【００３８】

また図９は、上記の文書ファイルに対して得られる、従来の索引ファイル３１の構成を示す図である。

上述したように、従来は、フィールドの位置情報をキーワードの位置情報と同様に索引ファイル３１へ登録していた。すなわち、キーファイル３２にフィールドを特定する情報を登録し、ＰＯＳファイル３３に当該フィールドの位置情報を登録していた。図９に示す例では、フィールド１の名称（フィールド１）がキーファイル３２に登録され、その位置情報（２，１）、（２，１２）がＰＯＳファイル３３に登録されている。キーワードに関しては、通常通り文字コード順にソートされ、重複するキーワードがまとめられて登録されている。

【００３９】

次に検索時の動作について説明する。

検索式が「検索ターム＠フィールド」という書式で記述されるものとし、「ＩＢＭ＠Ｆ１」というデータベース制御部２０から検索エンジン３０に送られたも

のとする。すなわち、「ＩＢＭ」という単語がフィールドＦ１に含まれている文書ファイルを検索する場合である。この場合、検索エンジン３０は、まず、関係ファイル３４を参照して、検索タームに該当するキーワード「ＩＢＭ」を得る。そして、このキーワード「ＩＢＭ」にフィールド１が組み合わされた登録「ＩＢＭ（Ｆ１）」があるかどうかを探す。

図８に示した索引ファイル３１のキーファイル３２には該当する登録「ＩＢＭ（Ｆ１）」が存在するので、ＰＯＳファイル３３に登録されている位置情報が直接参照される。これにより、位置情報（２，１）のみが検索結果として得られ、データベース制御部２０へ返される。

データベース制御部２０では、検索式「検索ターム＠フィールド」に対して得られた位置情報（２，１）に基づいて、文書番号２の文書ファイルを文書データベース１０から読み出すこととなる。

【００４０】

一方、図９に示した従来の索引ファイル３１を参照して検索する場合、検索タームに該当するキーワード「ＩＢＭ」からＰＯＳファイル３３に登録されている位置情報が参照される。同様に、検索式中のフィールドＦ１について、キーファイル３２からＰＯＳファイル３３に登録されているフィールドＦ１の情報が参照される。そして、これらの情報を突き合わせて、フィールドＦ１にキーワード「ＩＢＭ」が現れる文書が検索される。

具体的には、まず、キーワード「ＩＢＭ」の位置情報（２，１）に着目し、フィールドＦ１の位置情報と突き合わせる。フィールドＦ１は（２，１）から開始し、（２，１２）で終了するので、位置情報（２，１）のキーワード「ＩＢＭ」は検索式に適合する。次に、位置情報（２，２８）に着目すると、これはフィールドＦ１の定義から外れるので、位置情報（２，２８）のキーワード「ＩＢＭ」は検索式に適合しない。したがって、位置情報（２，１）のみが検索結果として得られ、データベース制御部２０へ返される。

【００４１】

以上のように、本実施の形態による検索エンジン３０は、図８に示したようにフィールド情報をキーワードに付随させて索引ファイル３１のキーファイル３２

及びＰＯＳファイル３３に登録している。検索式に該当する（すなわちフィールド情報も適合する）キーワードが得られたあとは、ＰＯＳファイル３３から当該キーワードの位置情報を取得するだけで良く、キーワードの位置情報とフィールドの位置情報とを突き合わせる作業を必要としない。したがって、図９に示した索引ファイル３１を用いる従来のシステムに比して、フィールド検索を含む検索処理に要する時間の大幅な短縮を見込むことができる。

【００４２】

なお、上記の検索動作では、フィールドごとに独立したキーワードに対応する関係ファイル３４を設定するのではなく、フィールド情報を含まないキーワードを関係ファイル３４で検索してから、フィールド情報と組み合わされたキーワードの登録を探した。したがって、フィールド情報と組み合わされたキーワードを独立にキーファイル３２に登録するとしても、関係ファイル３４の構成には影響を及ぼさず、検索処理に要する時間が増大することはない。

【００４３】

索引ファイル３１では、登録されているキーワードは文字コード順でソートされているので、フィールド情報と組み合わされたキーワードは、フィールド情報を含まない同一のキーワードの近傍に存在する。したがって、フィールド情報を含まないキーワードを得てから、キーファイル３２上でフィールド情報と組み合わされたキーワードを探すとしても、処理全体に影響を及ぼすほどの時間は要しない。

【００４４】

また、日本語の検索タームによる検索の場合のように、キーワードが固定長連鎖である場合は、関係ファイル３４を参照して検索タームに該当するキーワードを得る仕組みが存在しない。そのため、キーファイル３２の登録情報としてフィールド情報と組み合わされたキーワードが増えた分だけ、検索タームに該当するキーワードを探すのに要する時間が増加することとなる。しかしながら、キーファイル３２における登録データの増加分による処理の増加よりも、キーワードの位置情報とフィールドの位置情報とを突き合わせる作業を行わなくてすむことによる処理量の減少の方が、検索処理全体に与える影響が相当に大きいため、検索

処理の高速化に寄与すると考えられる。

【００４５】

なお、上記の実施の形態では、データベース検索システムを単体のコンピュータ装置にて実現する場合の構成を例として説明したが、本発明のシステムは、かかるハードウェア構成に限定されるものではない。文書データベース１０をネットワークで接続された他のコンピュータ装置において構築しても良いし、データベース制御部２０と検索エンジン３０とをネットワークで接続された別個のコンピュータ装置の機能として実現しても良い。

【００４６】

【発明の効果】

以上説明したように、本発明によれば、フィールド検索を含む文書データベースの検索において、高速な検索処理を実現することができる。

【図面の簡単な説明】

【図１】　本実施の形態におけるデータベース検索システムを実現するコンピュータ装置のハードウェア構成の例を模式的に示した図である。

【図２】　本実施の形態におけるデータベース検索システムの機能構成を示す図である。

【図３】　本実施の形態におけるデータベース制御部と検索エンジンとの間のデータのやり取りを示す図である。

【図４】　本実施の形態における索引ファイルの構成例を示す図である。

【図５】　本実施の形態における索引ファイルの他の構成例を示す図である。

【図６】　可変長連鎖をキーワードとする検索に用いられる関係ファイルの構成及びキーファイルとの関係を示す図である。

【図７】　本実施の形態による索引ファイルを用いたフィールド検索の方法を説明する図である。

【図８】　本実施の形態における索引ファイルのさらに他の構成例を示す図である。

【図９】　図８と同様の文書ファイルに対して得られる従来の索引ファイル

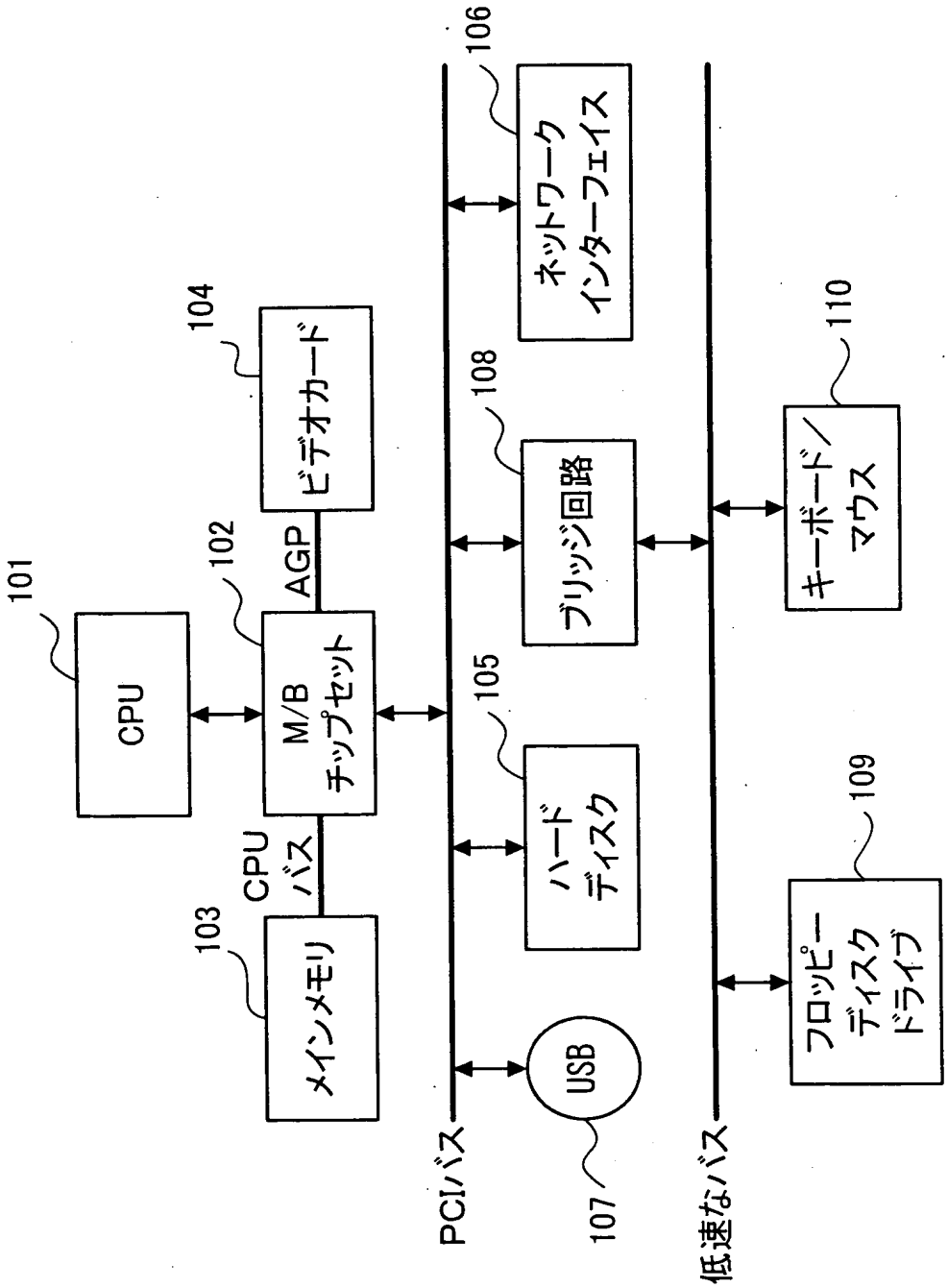
の構成を示す図である。

【符号の説明】

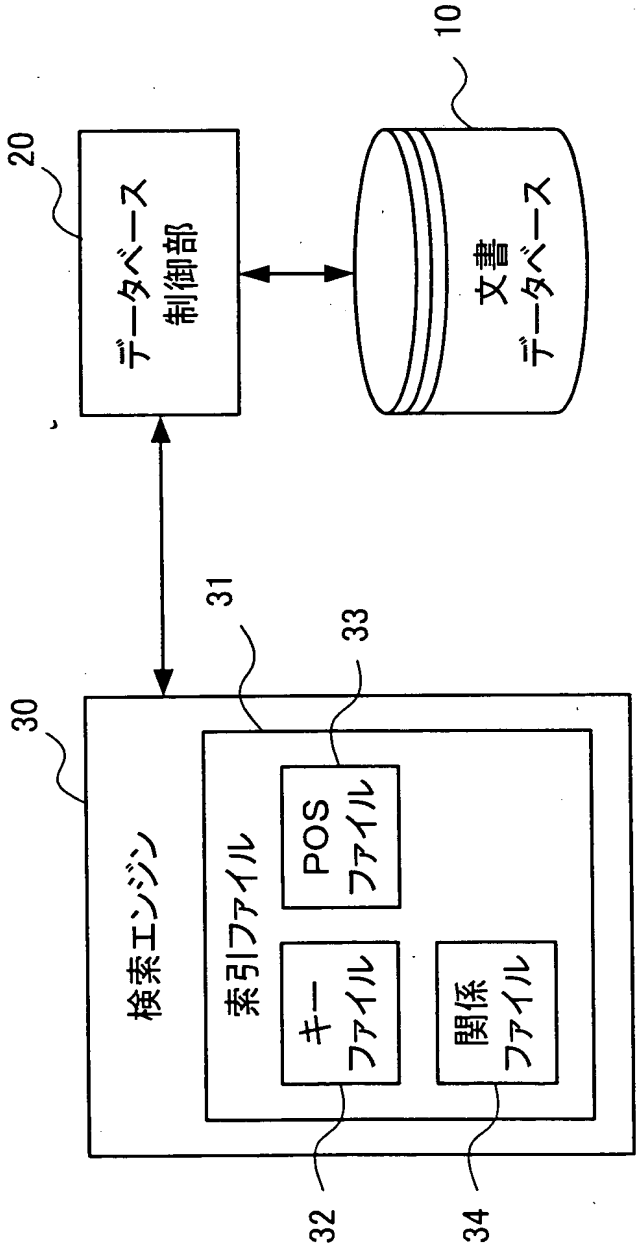
１０…文書データベース、２０…データベース制御部、３０…検索エンジン、３１…索引ファイル、３２…キーファイル、３３…ＰＯＳファイル、３４…関係ファイル、１０１…ＣＰＵ、１０３…メインメモリ、１０５…ハードディスク、１０６…ネットワークインターフェイス

【書類名】　　図面

【図１】

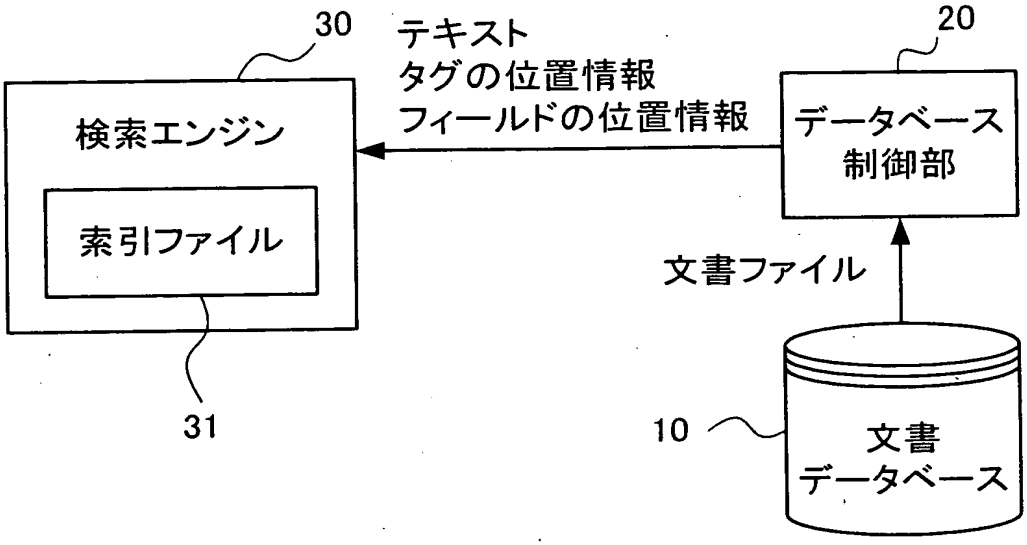

CPU
101
M/B
チップセット
102
AGP
ビデオカード
104
CPU
バス
メインメモリ
103
PCIバス
USB
107
ハード
ディスク
105
ブリッジ回路
108
ネットワーク
インターフェイス
106
低速なバス
キーボード／
マウス
110
フロッピー
ディスク
ドライブ
109

【図２】

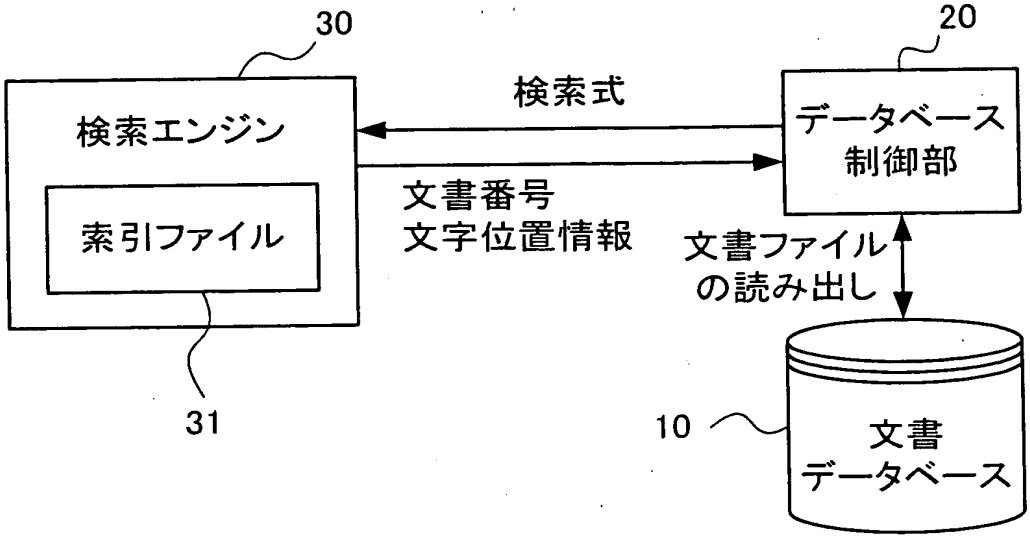

10
20
データベース
制御部
文書
データベース
31
33
30
検索エンジン
索引ファイル
POS
ファイル
キー
ファイル
関係
ファイル
32
34

【図３】

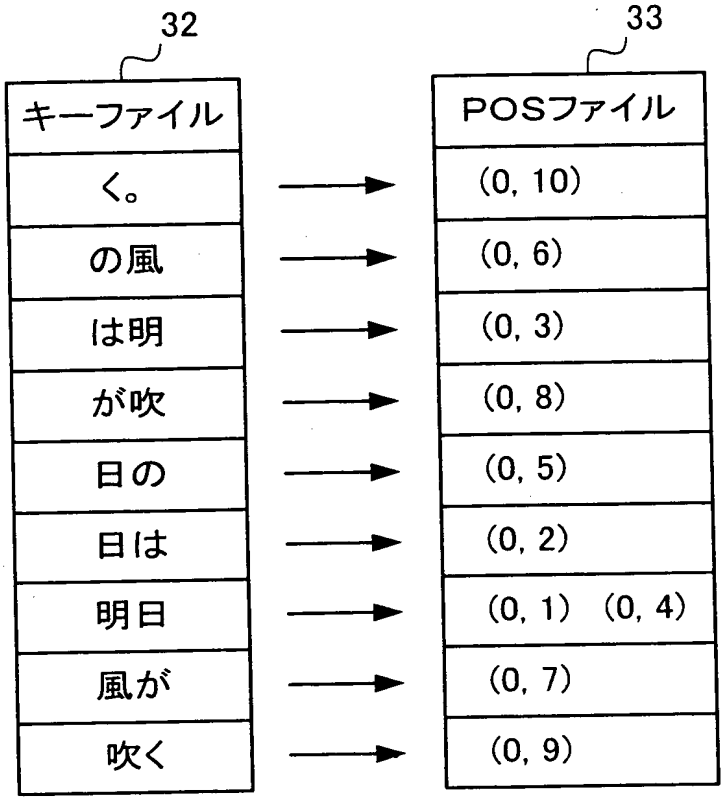
（A）


30
テキスト
タグの位置情報
フィールドの位置情報
20
検索エンジン
索引ファイル
データベース
制御部
文書ファイル
31
10
文書
データベース

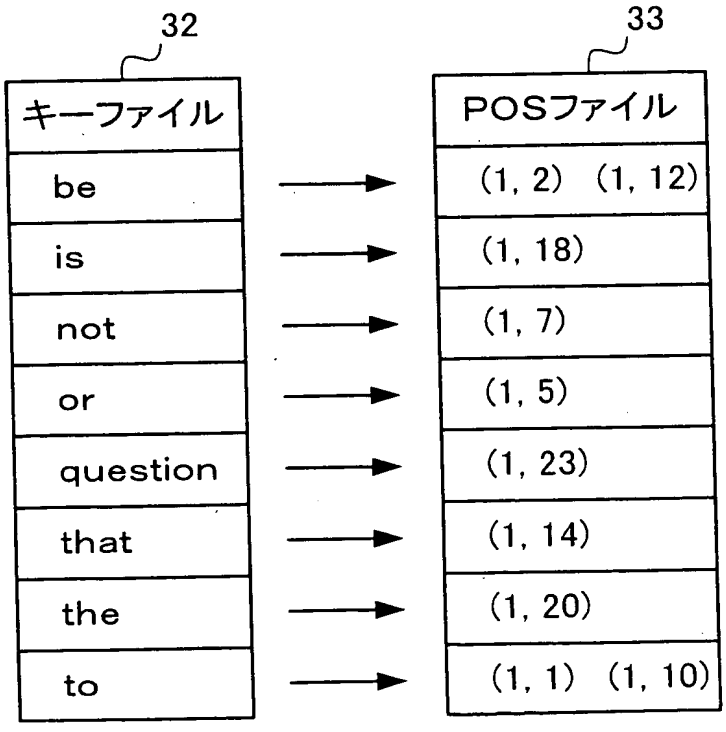
（B）


30
20
検索式
検索エンジン
データベース
制御部
索引ファイル
文書番号
文字位置情報
文書ファイル
の読み出し
31
10
文書
データベース

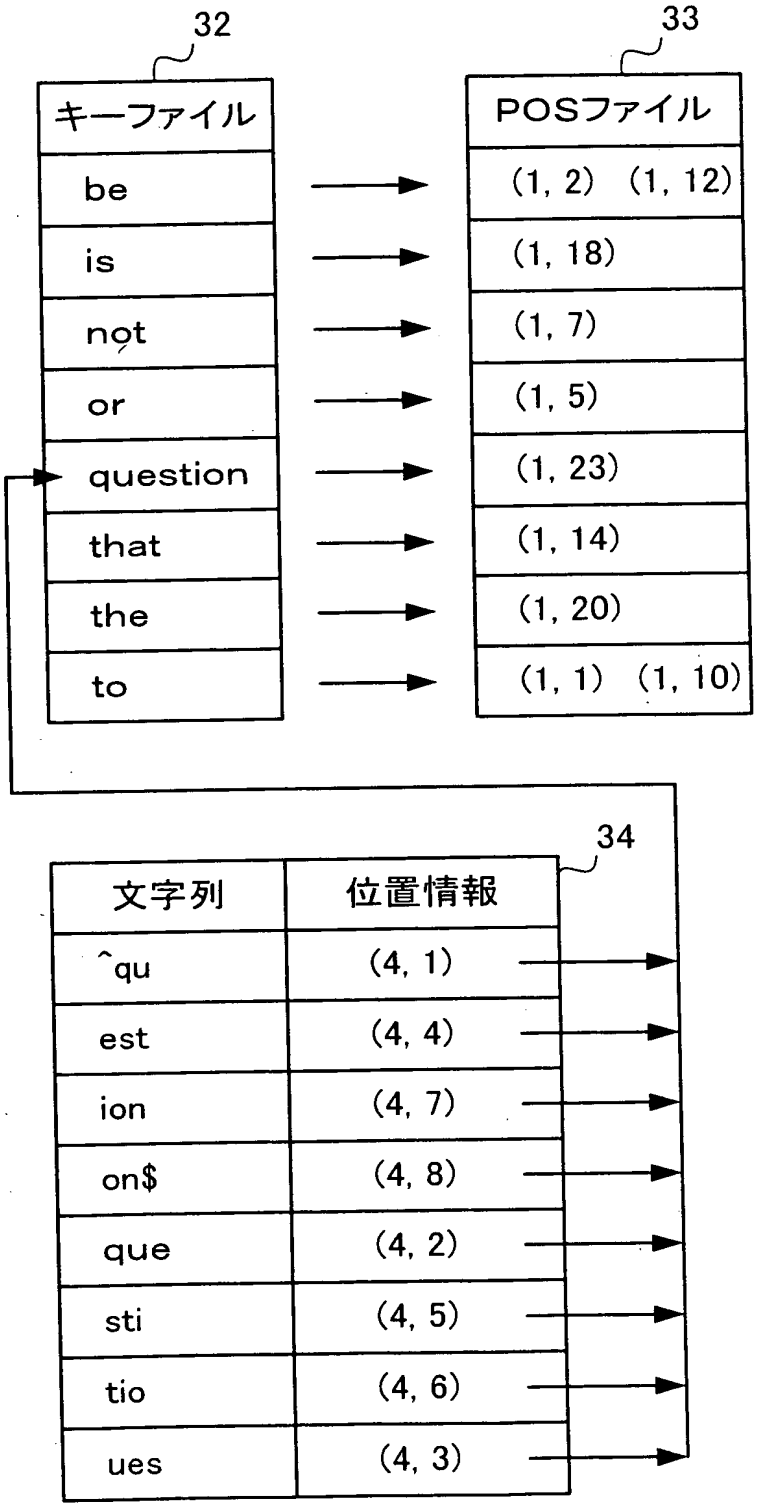
【図４】


32
33
キーファイル
POSファイル
く。
(0, 10)
の風
(0, 6)
は明
(0, 3)
が吹
(0, 8)
日の
(0, 5)
日は
(0, 2)
明日
(0, 1) (0, 4)
風が
(0, 7)
吹く
(0, 9)

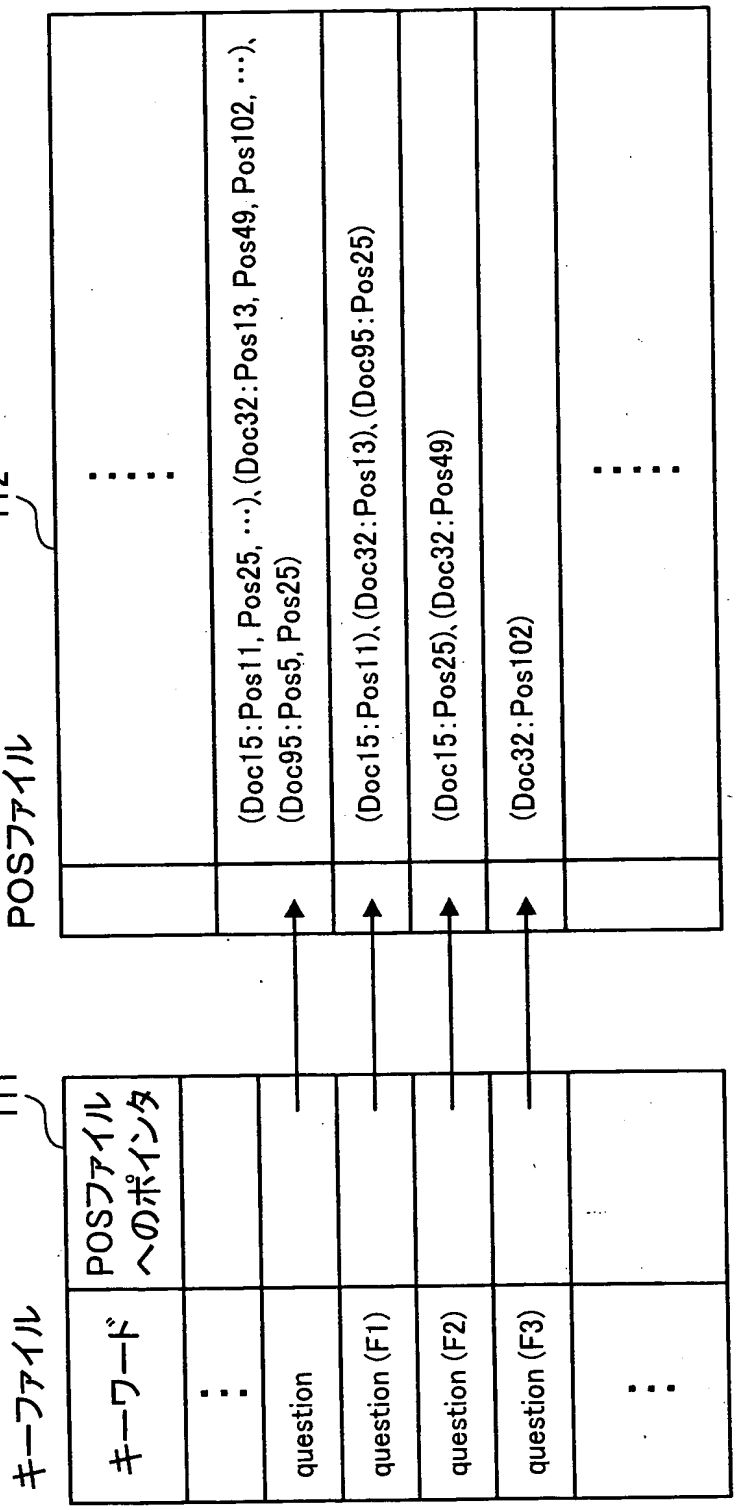
【図５】


32
33
キーファイル
POSファイル
be
(1, 2) (1, 12)
is
(1, 18)
not
(1, 7)
or
(1, 5)
question
(1, 23)
that
(1, 14)
the
(1, 20)
to
(1, 1) (1, 10)

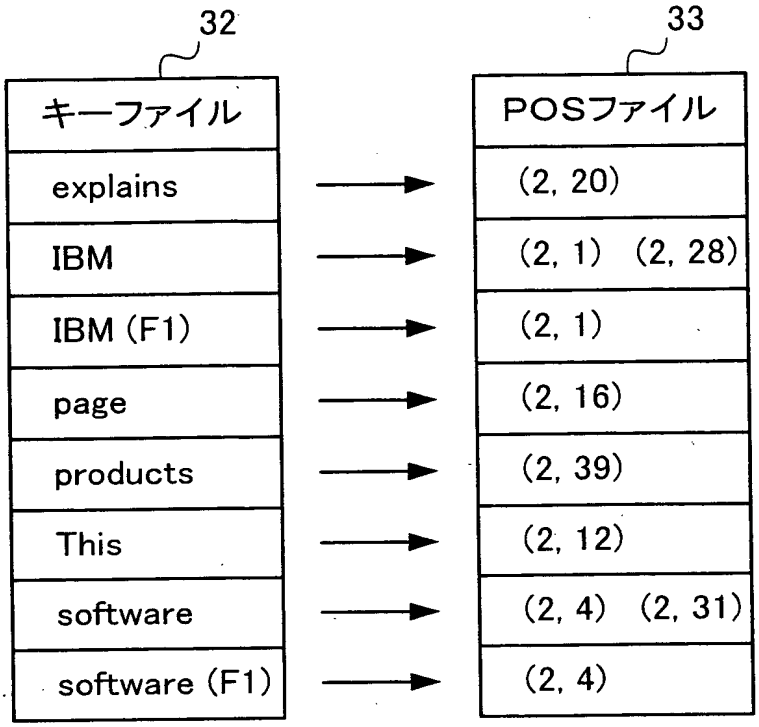
【図６】


32
33
キーファイル
POSファイル
be
(1, 2) (1, 12)
is
(1, 18)
not
(1, 7)
or
(1, 5)
question
(1, 23)
that
(1, 14)
the
(1, 20)
to
(1, 1) (1, 10)
34
文字列
位置情報
^qu
(4, 1)
est
(4, 4)
ion
(4, 7)
on$
(4, 8)
que
(4, 2)
sti
(4, 5)
tio
(4, 6)
ues
(4, 3)

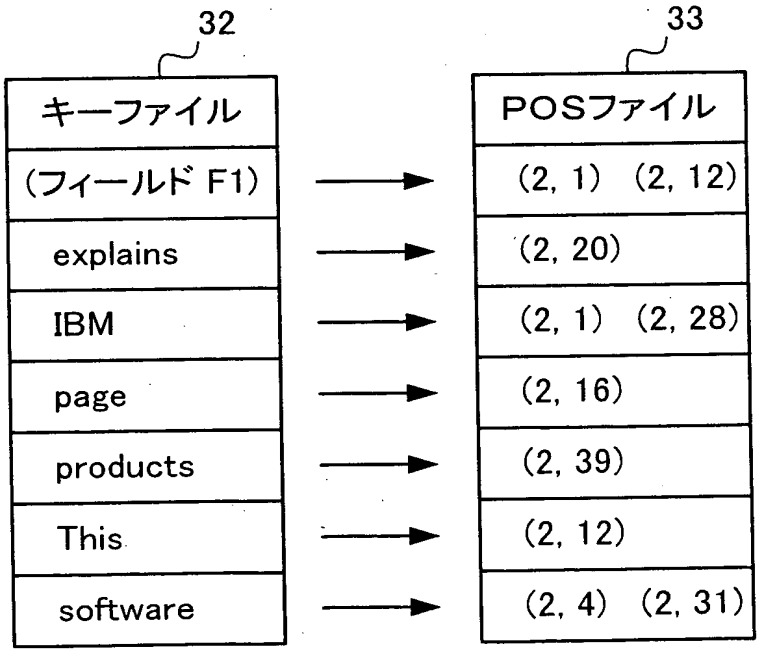
【図７】


キーファイル
111
キーワード
POSファイルへのポインタ
question
question (F1)
question (F2)
question (F3)
POSファイル
112
(Doc15:Pos11, Pos25, ・・・), (Doc32:Pos13, Pos49, Pos102, ・・・), (Doc95:Pos5, Pos25)
(Doc15:Pos11), (Doc32:Pos13), (Doc95:Pos25)
(Doc15:Pos25), (Doc32:Pos49)
(Doc32:Pos102)

【図８】


32
33
キーファイル
POSファイル
explains
(2, 20)
IBM
(2, 1) (2, 28)
IBM (F1)
(2, 1)
page
(2, 16)
products
(2, 39)
This
(2, 12)
software
(2, 4) (2, 31)
software (F1)
(2, 4)

【図９】


32
33
キーファイル
POSファイル
(フィールド F1)
(2, 1) (2, 12)
explains
(2, 20)
IBM
(2, 1) (2, 28)
page
(2, 16)
products
(2, 39)
This
(2, 12)
software
(2, 4) (2, 31)

【書類名】　　要約書

【要約】

【課題】　構造化された文書ファイルを蓄積した文書データベースに対する検索において、高速な検索処理を実現する。

【解決手段】　検索エンジン３０による検索処理に用いられ、キーワードとその位置情報との対応関係を示す情報を保持する索引ファイル３１を、文書データベース１０に蓄積された文書ファイルに含まれる文字列とこの文字列に関する位置情報へのポインタを、文字列が文書ファイル内の文字列が現れる文書領域別に登録したキーファイル３２と、このキーファイル３２に登録されている各文字列に関して、文字列が存在する文書ファイルを特定する情報及び文書ファイルにおける文字列の位置を特定する情報を含む位置情報を登録したＰＯＳファイル３３とを備える構成とする。

【選択図】　　図２

認定・付加情報

特許出願の番号	特願２００３－００４５７２
受付番号	５０３０００３４２９３
書類名	特許願
担当官	末武　実　　１９１２
作成日	平成１５年　２月２０日

＜認定情報・付加情報＞

【特許出願人】
- 【識別番号】 390009531
- 【住所又は居所】 アメリカ合衆国１０５０４、ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
- 【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】
- 【識別番号】 100086243
- 【住所又は居所】 神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　大和事業所内
- 【氏名又は名称】 坂口　博

【代理人】
- 【識別番号】 100091568
- 【住所又は居所】 神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　大和事業所内
- 【氏名又は名称】 市位　嘉宏

【代理人】
- 【識別番号】 100108501
- 【住所又は居所】 神奈川県大和市下鶴間１６２３番１４　日本アイ・ビー・エム株式会社　知的所有権
- 【氏名又は名称】 上野　剛史

【復代理人】 申請人
- 【識別番号】 100104880
- 【住所又は居所】 東京都港区赤坂５－４－１１　山口建設第２ビル６Ｆ　セリオ国際特許事務所
- 【氏名又は名称】 古部　次郎

【選任した復代理人】
- 【識別番号】 100118201

次頁有

認定・付加情報（続き）

【住所又は居所】　東京都港区赤坂５－４－１１　山口建設第二ビル
６Ｆ　セリオ国際特許事務所
【氏名又は名称】　千田　武

次頁無

出　願　人　履　歴　情　報

識別番号　〔３９０００９５３１〕

１．変更年月日　２００２年　６月　３日
［変更理由］　住所変更
住　所　アメリカ合衆国１０５０４、ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
氏　名　インターナショナル・ビジネス・マシーンズ・コーポレーション